UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/662,502	09/15/2003	Mu Li	M61.12-0527	9194

69316 7590 02/01/2010

MICROSOFT CORPORATION
ONE MICROSOFT WAY
REDMOND, WA 98052

EXAMINER	
SERROU, ABDELALI	
ART UNIT	PAPER NUMBER
2626	

NOTIFICATION DATE	DELIVERY MODE
02/01/2010	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

DBOUTON@MICROSOFT.COM
vffiling@microsoft.com
stevensp@microsoft.com

Office Action Summary	Application No.	Applicant(s)	
	10/662,502	LI ET AL.	
	Examiner	Art Unit	
	Abdelali Serrou	2626	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE *3* MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *04 November 2009*.

2a)☒ This action is **FINAL**. 2b)☐ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1-5,7-12,14-27 and 29-31* is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1-5, 7-12,14-27 and 29-31* is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☒ The specification is objected to by the Examiner.

10)☒ The drawing(s) filed on *15 September 2003* is/are: a)☒ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☐ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☐ Information Disclosure Statement(s) (PTO/SB/08) Paper No(s)/Mail Date _____
4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____.
5) ☐ Notice of Informal Patent Application
6) ☐ Other: _____.

DETAILED ACTION

Response to Amendment

1. In response to the office action mailed on 8/4/09, applicant filed an amendment on 11/4/09, amending claims 1, 14, and 25. Claims 6, 13, and 28 were previously cancelled. The pending claims are 1-5, 7-12, 14-27, and 29-31.

Response to Arguments

2. Applicant's arguments filed 11/4/09 have been fully considered but they are not persuasive.

As per claim 1, applicant argues that there is no disclosure within the prior art for using context including at least one Chinese character to determine the probability of a combination of character segments to be most probable. The examiner respectfully disagrees and point out that to determine the probability of a combination of character segments Brockett uses grammatical information, which is defined as the arrangement of words in sentences (context). Also, column 7, lines 38-40 states that Brockett's invention allows the normalized forms of any Chinese segment to be combined with other segments in the input string to identify a full segment for the input string of characters using context is necessarily disclosed within the teaching of Brockett. In order for Brockett to identify a sequence of characters and determine that this sequence of character forms a word, or verb, context has to be applied because the target character and the surrounding characters that form the word has to be considered. The system must take in consideration the characters combined together to determine whether they form a word or a verb as in the example of col. 6, lines 50-60, wherein a substring ABC is found in a text string and

rules, that necessarily include context, determine that the substring BC indicates the past tense for some verbs, and thereof the substring ABC is a verb in the past tense. For more see the description of Fig. 4 at Col. 6-7. Furthermore, applicant admits that Brockett teaches determining the highest word probability (see Remarks, page 10, line 22. See also, col. 6, wherein probability is used to determine possible words from text strings). Therefore, the probability used by Brockett is related to context.

Applicant argues that Brockett does not disclose utilizing forward and backward maximum matching searches. The examiner notes that this feature is taught by the primary reference Chen

As per claims 14 and 25, applicant argues that the claims determine constituent lexical words in the overlapping ambiguity string instead of left and right portions in claim 1. Applicant did not explain how the right/left portions are different from constituent lexical words in the overlapping ambiguity string. According to the specification, the constituent lexical words in the overlapping ambiguity string represent the right/left portions of the overlapping ambiguity string. As to utilizing an n-gram model to obtain probability, the prior art Brockett uses n-gram models (col. 2, lines 47-48).

As per the rest of the claims, and combinations of prior art reference, applicant has no further arguments beside the ones mentioned above. Therefore, all the combinations of prior art reference mentioned above are valid, and all other claims are rejected for the same reasons as set above.

Specification

3. The disclosure is objected to because of the following informalities: paragraph [0072]: step 716 of Fig. 7 should be step 706. Appropriate correction is required.

Claim Objections

4. Claim 1 is objected to because of the following informalities: lines 16-18 of claim 1, recite "wherein the probability information is based on at least one context feature adjacent one of the right portion or left portion of each of the possible segmentation". The examiner interprets the above limitation as "wherein the probability information is based on at least one context feature adjacent **to** one of the right portion or left portion of each of the possible segmentation". Appropriate correction is required.

Claim Rejections - 35 USC § 103

5. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

Claims 1-4, 7, 14, 15-21, 23, 25-26 are rejected under 35 U.S.C. 103(a) as being unpatentable over Chen et al. (U.S 5,806,021 issued on Sept. 8, 1998) (hereinafter: Chen) in view of Brockett et al. (U.S 6,968,308, filed Nov. 1, 2000 and issued on Nov. 22, 2005) (hereinafter: Brockett).

As per claims 1, 14, and 25, Chen teaches segmenting a sentence of Chinese characters into constituent Chinese words having one or more Chinese characters by performing a Forward Maximum Matching (FMM) segmentation of the input sentence and a Backward Maximum Matching (BMM) segmentation of the input sentence to generate a first and second set of tokens (col. 3, lines 18-32, wherein a Forward and Backward Maximum Matching segmentations are performed); generating an n-gram model (col. 4, lines 45-47), and selecting one of the two segmentations as a function of probability information for the two segmentations (col. 4, lines 25-26); and outputting an indication for selecting one of the at least two possible segmentations as a function of the obtained probability information (col. 3, lines 29-32, wherein the likelihood of the segmentation is calculated and the one with the higher likelihood is chosen as a result); and outputting an indication for selecting one of the at least two possible segmentations, FMM and BMM segmentation, as a function of the probability information (col. 3, lines 18-32, wherein segmentations that correspond to both directions, forward and backward, are obtained and the one with higher probability is chosen).

Chen does not explicitly teach tokenizing the sentence into common tokens and differing tokens for recognizing an overlapping ambiguity string in the segmented sentence, wherein the overlapping ambiguity string comprises at least three Chinese characters (constituent lexical words) having at least two possible segmentations wherein each possible segmentation comprises a right portion and a left portion and wherein the right portion and left portion of each of the possible segmentations (constituent lexical words) remains in a tokenized corpus and at least the overlapping ambiguity string is removed from the tokenized corpus, and obtaining probability information related to context for each possible segmentation of the at least three

Chinese characters, wherein the probability information is based on at least one context feature adjacent the overlapping ambiguity string and one of the right portion or left portion of the possible segmentation, and wherein the at least one context feature comprises a Chinese character.

Brockett in the same field of endeavor teaches tokenizing the sentence into common tokens and differing tokens for recognizing the overlapping ambiguity string in the segmented sentence, wherein the overlapping ambiguity string comprises at least three Chinese characters (constituent lexical words) having at least two possible segmentations with right and left portions and wherein the right portion and left portion remain (constituent lexical words) in a tokenized corpus and at least the overlapping ambiguity string is removed from the tokenized corpus, (col. 1, lines 40-48, wherein the processed text is non-segmented text like Japanese or Chinese; col. 2, lines 16-17 and col. 10, lines 41-49, wherein the recognized overlapping ambiguity string comprises at least three Chinese characters having at least two possible segmentations. As an example: a sentence represented by characters ABCD. There are at least two possible segmentations, A/BCD, AB/CD, and ABC/D; and for a sentence represented by characters ABC, A/BC and AB/C would be the possible segmentations. The overlapping ambiguity string comprises at least three Chinese characters or constituent lexical words. Each possible segmentation has a left portion, wherein the right portion and left portion remain (constituent lexical words) in a tokenized corpus, and the overlapping ambiguity string is removed from the tokenized corpus), obtaining probability information related to context based on at least one context feature adjacent to one of the right portion or left portion of each of the possible segmentation the overlapping ambiguity string and at least part of the recognized OAS for each

of the FMM and BMM (necessarily disclosed within the process of col. 6, lines 6-42, wherein the system checks the context feature of adjacent to the OAS to identify substrings AB, BC, ABC of the string ABCD); and replacing the overlapping ambiguity string with tokens (necessarily disclosed in selecting the most probable segmentation for the input string (col. 11, lines 5-19). As to utilizing an n-gram model to obtain probability, the prior art Brockett uses n-gram models (col. 2, lines 47-48). As to determining at least two different pairs of constituent lexical words in the overlapping ambiguity string, is necessarily disclosed for an overlapping ambiguity string represented by at least three Chinese characters ABC, the constituent lexical words could be A/ BC and AB/ C.

Therefore, it would have been obvious to a person of ordinary skill in the art at the time of the invention was made to apply the features of the overlapping ambiguity string recognizer of Brockett to the text segmentation system of Chen, to resolve the overlapping ambiguity of unsegmented input strings, because Brockett suggests that this would better identify the right segment among the competing segments (col. 1, lines 55-63).

As per claims 2-4, 23, and 26, Chen in view of Brockett teach obtaining the probability information from a language model (lexicon, col. 2, line 41) based on the at least one context feature and a left or right portion of the overlapping ambiguity string (necessarily disclosed for determining word boundaries, col. 2, lines 39-44), wherein the language model comprises a trigram model (col. 2, lines 45-49), wherein outputting an indication for selecting one of the at least two possible segmentations comprises classifying the probability information (col. 3, lines 29-32, wherein the probability information (likelihood) of both segmentations is calculated and classified to select the segmentation with higher likelihood).

As per claim 7, Chen teaches performing a Forward Maximum Matching (FMM) segmentation, for recognizing a segmentation O_f, (col. 3, lines 15-65) and a Backward Maximum Matching (BMM) segmentation for recognizing a segmentation O_b of the input sentence (col. 3, line 15 - col. 4, line 24).

Chen does not explicitly teach recognizing an overlapping ambiguity string in the input sentence as a function of the two segmentations.

Brockett in the same field of endeavor teaches recognizing the overlapping ambiguity string in the input sentence as a function of the two segmentations (col. 2, lines 16-17).

Therefore, it would have been obvious to a person of ordinary skill in the art at the time of the invention was made to combine the overlapping ambiguity string recognizer of Brockett to the text segmentation system of Chen, because Brockett suggests that this would better identify the right segment among the competing segments (col. 1, lines 55-63).

As per claim 15, Chen teaches determining a probability associated with each of the FMM segmentation of the overlapping ambiguity string and the BMM segmentation of the overlapping ambiguity string based on higher probability (col. 3, lines 18-32, wherein the segmentation with higher likelihood is chosen).

As per claims 16-18, Chen teaches an N-gram model (col. 4, lines 45-47), and probability information about a first and last word of the overlapping ambiguity string (col. 5, lines 1-5, wherein probability of each part of the phrase (word), resulted from a segmentation is compared separately).

As per claims 19-21, Chen teaches N-gram model (col. 4, lines 45-47), that uses trigram probability information about a string of words comprising a first word of the overlapping

ambiguity string and two context words to the left of the first word, and a last word of the overlapping ambiguity string and two context words to the right of the last word (inherently disclosed in the process of determining likelihood scores using n-grams models (tri-gram model), col. 5, lines 45-47).

Claims 5, 8-12, 22, 24, 27, and 29-31, are rejected under 35 U.S.C. 103(a) as being unpatentable over Chen in view of Brockett, as applied to claims 4, 15, and 23, and further in view of Pedersen ("*A Simple Approach to Building Ensembles of Naive Bayesian Classifiers for Word Sense Disambiguation*", in Proceedings of the First Annual Meeting of the North American Chapter of the Association for Computational Linguistics, pp. 63-69, April 29 – May 4, 2000).

As per claim 5, 22, and 24, Chen in view of Brockett teaches all the limitations of claims 4, 15, and 23, upon which claims 5, 22, and 24 depend.

Chen and Brockett do not explicitly teach using an ensemble of Naive Bayesian Classifiers.

Pederson in the same field of endeavor teaches using an ensemble of Naive Bayesian Classifiers (Abstract).

Therefore, it would have been obvious to a person of ordinary skill in the art at the time of the invention was made to combine Pederson's Nave Bayesian Classifier with the automatic text segmenter of Chen, because Pederson suggests that this would provide more accurate disambiguation systems (Abstract).

As per claims 8-12, Chen in view of Brockett teach one of the two segmentations (col. 4, lines 25-26), classifying the probability information of O_f and O_b (col. 3, lines 29-32, wherein

the probability information (likelihood) of both segmentations is calculated and classified to select the segmentation with higher likelihood), and determining which one of the said probabilities is higher (col. 4, lines 25-26).

Chen and Brockett do not explicitly selecting one of the at least two segmentations as a function of a set of context features, words around the overlapping ambiguity string, associated with the overlapping ambiguity string, classifying the probability information of the context features surrounding the overlapping ambiguity string, and determining which one of the said probabilities is higher, as a function of the set of context features.

Pederson in the same field of endeavor teaches the Naïve Bayesian Classifier for word sense disambiguation based on windows of context (Pages 63-64).

Therefore, it would have been obvious to a person of ordinary skill in the art at the time of the invention was made to use the Naïve Bayesian Classifier of Pederson in combination with the text segmenting system of Chen, to use the probability information of the context features to select one of the two segmentations. Pederson suggests that this would provide more accurate disambiguation systems (Abstract).

As per claims 27 and 29, Chen in view of Brockett teaches all the limitations of claims 25 and 28, upon which claims 27 and 29 depend.

Chen and Brockett do not explicitly teach generating an ensemble of classifiers as a function of an n-gram model.

Pederson in the same field of endeavor teaches generating an ensemble of classifiers as a function of an n-gram model (Abstract, and page 64, col. 2, lines 15-19).

Therefore, it would have been obvious to a person of ordinary skill in the art at the time of the invention was made to combine Pederson's classifiers with the combined system of Chen and Brockett, because Pederson suggests that this would provide more accurate disambiguation systems (Abstract).

As per claim 30, Chen, Brockett, and Pederson teach all the limitations of claim 29, upon which claim 30 depends. Chen in view of Brockett, furthermore, teach approximating probabilities of the FMM and BMM segmentations of each overlapping ambiguity string as being equal to the product of individual unigram probabilities of individual words in the FMM and BMM segmentations respectively, of the overlapping ambiguity string (col. 3, line 37 –col. 4, line 26, wherein the probabilities of the FMM and BMM segmentations of each overlapping ambiguity are approximated and compare to choose the one with the highest score).

As per claim 31, Chen, Brockett, and Pederson teach all the limitations of claim 30, upon which claim 31 depends. Pederson, furthermore, teach a joint probability of a set of context features conditioned on an existence of one of the segmentations of each overlapping ambiguity string (ambiguous word) as a function of a corresponding probability of a leftmost and a rightmost word of the corresponding overlapping ambiguity string (Pages 63-64, 2nd paragraph, NaiveBayesian Classifiers).

Conclusion

Examiner has cited particular columns and line numbers in the references applied to the claims above for the convenience of the applicant. Although the specified citations are representative of the teachings of the art and are applied to specific limitations within the individual claim, other passages and figures may apply as well. It is respectfully requested from

the applicant in preparing responses, to fully consider the references in entirety as potentially teaching all or part of the claimed invention, as well as the context of the passage as taught by the prior art or disclosed by the Examiner.

In the case of amending the claimed invention, Applicant is respectfully requested to indicate the portion(s) of the specification which dictate(s) the structure relied on for proper interpretation and also to verify and ascertain the metes and bounds of the claimed invention.

When responding to this office action, applicants are advised to clearly point out the patentable novelty which they think the claims present in view of the state of the art disclosed by the references cited or the objections made. Applicants must also show how the amendments avoid such references or objections. See 37C.F.R 1.111(c). In addition, applicants are advised to provide the examiner with the line numbers and pages numbers in the application and/or references cited to assist examiner in locating the appropriate paragraphs.

THIS ACTION IS MADE FINAL. Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire THREE MONTHS from the mailing date of this action. In the event a first reply is filed within TWO MONTHS of the mailing date of this final action and the advisory action is not mailed until after the end of the THREE-MONTH shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the mailing date of this final action.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Abdelali Serrou whose telephone number is 571-272-7638. The examiner can normally be reached on 8:30-5:00.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, David R. Hudspeth can be reached on 571-272-7843. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

/David R Hudspeth/
Supervisory Patent Examiner, Art Unit 2626